

ПРОЦЕДУРА ПОСТРОЕНИЯ ЧАСТОТНОГО СЛОВАРЯ НА ОСНОВЕ ЛЕКСИЧЕСКИ СВЯЗАННЫХ КОМПОНЕНТОВ

Рассмотрена процедура построения частотного словаря на основе лексически связанных компонентов и практические аспекты ее применения. Сформулирована задача о перестройке частотного мультилингвистического словаря для информационного обеспечения методики обучения иностранной лексики на основе лексически связанных компонентов.

Ключевые слова: мультилингвистическая адаптивно-обучающая технология, лексически связанный компонент, информационно-терминологический базис, ЛСК-методика.

Как бы стремительно ни развивалась наука и техника, основу обучения иностранной терминологии всегда будут составлять словари. Они видоизменяются, переключаются на компьютерные системы, делятся по тематикам, но по-прежнему представляют собой не что иное, как совокупность терминов и их языковых аналогов. Что касается специализированных словарей, таких, например, как словарь по медицине или математическому моделированию, то они могут быть построены с помощью программных систем, анализирующих тексты. В основе алгоритмизации этих систем, как правило, лежит частотный анализ терминов.

Словарь, полученный таким образом, перед выходом в печать проходит контроль со стороны лингвистов и специалистов конкретных предметных областей, и чем более качественно реализована программная часть, тем более поверхностный характер имеет их работа. Одним из таких словарей является изданный в 2004 г. англо-немецко-русский частотный словарь по системному анализу [1]. Этот словарь стал одним из результатов развития мультилингвистической адаптивно-обучающей технологии (МЛ-технологии) [2], а если говорить точнее, то он является одной из форм ее информационно-терминологического базиса (ИТБ).

Мультилингвистичность словаря делает возможным в процессе его использования построение однозначных ассоциативных связей между языковыми аналогами английского, немецкого и русского языка при условии, что обучаемый или уже знает один из иностранных языков, или изучает два иностранных языка одновременно. Здесь и далее русский язык рассматривается в качестве оригинального.

Дальнейшее развитие МЛ-технологии привело к созданию методики обучения иностранной лексике на основе лексически связанных компонентов (ЛСК-методике) [3], которая позволяет искусственно формировать строго организованные системы внутриязыковых ассоциативных связей непосредственно в процессе обучения иностранной лексике. Эта методика также учитывает языковые аналоги лексем на всем множестве языков, с которыми она работает.

ЛСК-методика использует специально организованные ИТБ. Их формирование осуществляется посредством анализа текстов на предмет устойчивых лексических сочетаний.

Согласно ЛСК-методике, ИТБ представляет собой совокупность лексически связанных компонентов (ЛС-ком-

понентов). Лексема, связанная со всеми без исключения лексемами ЛС-компонента ИТБ, называется *основной лексемой*, а лексемы, имеющие только одну связь, – *связанными лексемами*.

В данной статье речь пойдет о решении реальной задачи перестройки частотного мультилингвистического словаря для использования ЛСК-методики. Решение этой задачи целесообразно разделить на следующие этапы:

- подготовка необходимых данных для применения алгоритмов формирования ЛС-компонентов;
- выбор и применение алгоритма формирования ЛС-компонентов;
- корректировка результатов.

Остановимся на каждом этапе более подробно.

Получение необходимых данных для применения алгоритмов формирования ЛС-компонентов. Для использования этих алгоритмов необходимы следующие данные: абсолютные и относительные частоты лексем, данные о лексических связях, в том числе их частотные характеристики.

До обработки словарь представляет собой таблицу, состоящую из терминов английского, немецкого и русского языков, в которой также определены частоты терминов (рис. 1).

Применение алгоритмов формирования ЛС-компонентов основано на том, что каждый элемент словаря должен содержать полную информацию о лексических связях в виде ссылки на связанные лексемы и частоты лексических сочетаний. Иначе говоря, информационный базисный компонент словаря необходимо привести к следующему виду:

$$\text{ОМЛ-компонент} = \{ \text{термин яз}_1, \text{термин яз}_2, \dots, \text{термин яз}_N, \text{частота яз}_1, \text{частота яз}_2, \dots, \text{частота яз}_N, \text{сочетание } 1_яз_1, \text{сочетание } 1_яз_2, \dots, \text{сочетание } 2_яз_1, \text{сочетание } 2_яз_2, \dots, \text{сочетание } K_яз_N, \text{частота } 1_яз_1, \text{частота } 1_яз_2, \dots, \text{частота } 2_яз_1, \text{частота } 2_яз_2, \dots, \text{частота } K_яз_N \}.$$

Это возможно путем прогона использованных при формировании словаря текстов через новые алгоритмы обработки. Но поскольку результаты программной обработки текстов при формировании первой версии словаря корректировались специалистами, то такой подход не даст точных результатов, а повторная экспертиза иностранной терминологии – задача достаточно трудоемкая и дорогостоящая и также не может гарантировать необ-

ходимой точности, если речь не идет о формировании словаря заново.

Вместе с тем необходимые данные можно получить без дополнительных затрат, проанализировав терминологический состав словаря. Здесь следует пояснить, что устойчивые лексические сочетания, которые мы ищем, отражают отдельные понятия языка и наиболее востребованные из них вошли в состав словаря как отдельные элементы со своими частотами. Узкое место такого подхода состоит в том, что далеко не все термины, составляющие лексические сочетания, присутствуют в словаре. Однако частота употребления этих терминов равна или, скорее всего, превосходит частоту сочетаний, в которых они употребляются. Это дает нам возможность пополнить словарь недостающими терминами и их языковыми аналогами.

В рассматриваемом нами случае за основу взяты английские термины, затем был выполнен их адекватный перевод на другие языки и определены соответствующие частоты (рис. 2).

Выбор и применение алгоритма формирования ЛС-компонентов. Поскольку объем базиса невелик, а требования к количеству ЛС-компонентов отсутствуют, то целесообразно применить восходящий алгоритм формирования ЛС-компонентов (В-алгоритм) [3]. В нем можно выделить следующие фазы.

1. Подготовка ИТБ.

1.1. Для каждой лексемы ИТБ вычисляется значение $L_i, i = 1, \dots, n$:

$$L_i = e^{\frac{0,7}{\sum_k q_k \mu_{ik} + 1}} q_i \rightarrow \max, \quad (1)$$

где μ_{ik} – относительная частота сочетания i -й и k -й лексем, отражающая силу ассоциативной связи; q_i – относительная частота, выражающая долю лексической единицы в тексте, подвергшемся статистической обработке при составлении частотного словаря, $0 < q_i < 1$:

$$q_i = \frac{q_i^{\max}}{V}, \quad (2)$$

здесь $q_i^{\max} = \max q \{q_{i1}, q_{i2}, \dots, q_{in}\}$ – абсолютная частота появления лексической единицы в тексте, $q_{i1}, q_{i2}, \dots, q_{in}$ – частоты из мультилингвистического словаря, если речь идет о МЛ-технологии.

1.2. ИТБ упорядочивается по убыванию значения L_i таким образом, что чем меньше будет порядковый номер лексемы, тем выше вероятность образования на ее основе ЛС-компонента.

2. Поиск оптимального количества основных лексем.

2.1. Осуществляется перебор возможного количества основных лексем k от 1 до объема ИТБ (возможно сужение разработчиком интервала поиска).

2.2. Для текущего значения k определяются основные лексемы (k первых лексем ИТБ).

2.3. Осуществляется перебор неосновных (потенциально связанных) лексем и для каждой неосновной лексемы из множества, сформированного в п. 2.2, выбирается наиболее подходящая основная лексема согласно критерию

$$\Delta L_i = q_i e^{\frac{0,7}{(\sum_k q_k \mu_{ik} + 1) + q_j \mu_{ij}}} - L_i \rightarrow \max. \quad (3)$$

Таким образом вычисляется максимальный прирост $L(n)$, который обеспечивается вхождением текущей неосновной j -й лексемы как связанной в ЛС-компонент, образованный i -й лексемой (как правило, задается максимум количества связанных лексем).

2.4. Подсчитывается значение функции качества:

$$L(n) = \sum_i q_i e^{\frac{0,7}{\sum_k q_k \mu_{ik} + 1}} \rightarrow \max, \quad (4)$$

которая показывает сумму взвешенных вероятностей знания лексем по всему базису. Чем больше эта сумма, тем более удачно построен базис.

	1	2	3	4
1	1 abbreviate, 8	abkürzen, 8	abkürzen, 8	сокращать, 11
2	2 abbreviate a notation, 3	ein Bezeichnung abkürzen, 3	ein Bezeichnung abkürzen, 3	сокращать обозначение, 3
3	3 ability, 4	Fähigkeit, 4	Fähigkeit, 4	способность, 5
4	4 above, 32	über, 30	über, 30	сверх, свыше, 18
5	5 abstraction, 7	Abstraktion, 2	Abstraktion, 2	абстракция, 42
6	6 actual, 20	aktuell, 17	aktuell, 17	актуальность, 18
7	7 acceptance, 5	Annahme, 5	Annahme, 5	понятие, 33, прием, 31
8	8 access, 49	Zugriff, 42	Zugriff, 42	доступ, 25, взаимодействие, 37
9	9 direct access, 2	direkter Zugriff, 3	direkter Zugriff, 3	прямой способ доступа, прямой доступ
10	10 display access, 3	Displayzugriff, 2	Displayzugriff, 2	обращение к дисплею, 5
11	11 double junction, 2	unabhängiger Zugriff, 4	unabhängiger Zugriff, 4	двойной доступ, 2
12	12 parallel access, 3	paralleler Zugriff, 3	paralleler Zugriff, 3	параллельная выборка, 2
13	13 sequential access, 2	sequenzieller Zugriff, 3	sequenzieller Zugriff, 3	последовательная выборка, 3
14	14 random access, 3	Zufallszugriff, 4	Zufallszugriff, 4	случайная выборка, 2
15	15 simultaneous access, 4	gleichzeitiger Zugriff, 2	gleichzeitiger Zugriff, 2	одновременная выборка, 2
16	16 storage access, 5	Speicherzugriff, 2	Speicherzugriff, 2	доступ к памяти, доступ к устройству
17	17 access method, 6	Zugriffsmethode, 4	Zugriffsmethode, 4	метод доступа, 3
18	18 accessibility, 2	Zugänglichkeit, 3	Zugänglichkeit, 3	уровень доступа, 2
19	19 accessibility, 4	Zugänglichkeit, 2	Zugänglichkeit, 2	доступность, 3

Рис. 1. Частотный словарь до обработки

2.5. Если перебор окончен, то выполняется п. 2.6, если иначе, то возврат к п. 2.1.

2.6. Определяется максимум функции качества – оптимальное число основных лексем k_{max} .

3. Формирование ИТБ как совокупности ЛС-компонентов. Искомый ИТБ получается при прохождении пп. 2.2 и 2.3 для k_{max} основных лексем.

С помощью В-алгоритма было выявлено и сформировано 70 ЛС-компонентов (рис. 3).

Корректировка результатов и оформление словаря. Таким образом, мы получили двухблочный словарь: первый блок состоит из ЛС-компонентов, второй – из элементов, которые не вошли в первый блок. Для удобства использования словаря следует выполнить сортировку блоков по алфавиту (ЛС-компоненты упорядочиваются по основным лексемам) и добавить соответствующие литеры.

Если словарь разработан для печати, то рационально добавить во второй блок терминологию первого блока со ссылками-номерами на соответствующие ЛС-компоненты. В таком случае по второму блоку пользователь может производить поиск всей интересующей его терминологии, а по первому обучаться согласно процедурам ЛСК-методики.

Также рекомендуется провести корректировку терминологии в ручном режиме.

Таким образом, рассмотрена процедура построения частотного словаря на основе лексически связанных компонентов и практические аспекты ее применения. В качестве основы взята реальная задача о перестройке частотного мультилингвистического словаря для информационного обеспечения ЛСК-методики. Даны рекоменда-

№	1	2	3	4	5	6	7	8
7	1	access 49	Zugang 48 m	доступ, 26	50b	49	der Zugang, der Zugang, m	1) вход; 2) доступ
8	3	accessible 2	Zugänglich 3 f	доступ, 26	50b	3		
9	5	accessibility 4	Zugänglichkeit 2	доступ, 26	50b	4		
10	10	accessory 13	Hilfsausrüstung 2	вспомогательное		13		
11	14	accuse 14	verurteilen 4	обвинять, 28		14		
12	12	accumulate 3	ansammeln (syn. aufhäufen) 3, 4	накапливать, 3		31		
13	12	accumulation 13	Akkumulation 3 f	накапливание, инфр		13	act of accum. & accumulation; пункт накопления, пункт	
14	14	accumulator 59	Akkumulator 32 m	накопитель, 59a		59	der Akkumulator, der Akkumulator, m	накопитель, 59a
15	15	accuracy 17	Genauigkeit 35 f	точность, 35		35	die Genauigkeit, die Genauigkeit, f	точность, 35
16	16	accuracy 17	Genauigkeit 35 f	точность, 35		35	die Genauigkeit, die Genauigkeit, f	точность, 35
17	17	acquisition 5	Erwerb 12 m	приобретение, 12		12	der Erwerb, der Erwerb, m	приобретение, 12
18	18	acq. 164	Erwerb 12 m	приобретение, 12		12	der Erwerb, der Erwerb, m	приобретение, 12
19	18	activate 9	aktivieren 7	активировать, 7		25		
20	20	active 11	aktiv 4	активный, 11		11		
21	21	activity 10	Aktivität 4 f	деятельность, 10		10	die Aktivität, die Aktivität, f	деятельность, 10
22	22	accept 12	annehmen 14	принимать, 14		17		
23	23	acceptably 5	Annehmbarkeit 35 f	пригодность, при		5		
24	24	adaptation 4	Anpassung 4 f	адаптация, 4		4		
25	25	accepted 6	angenommen 6	принятый, 6		7		
26	26	acceptive 14	adaptiv 7	адаптивный, 7		14		
27	27	acc 25	akkumulieren 3	накапливать, 3		45		

Рис. 2. Словарь после этапа подготовки данных

№	1	2	3	4	5
1	А	А			
2	+	1 accuracy 47	Genauigkeit 35 f		точность, 35
3		2 accuracy 31	Genauigkeit 35 f		точность, 31
4		3 accuracy accuracy 8	Genauigkeit 35 f		точность, 8
5		4 available 18	verfügbar 2		доступный, 18
6		5 available accuracy 3	verfügbare Genauigkeit 3 f		доступная точность, 3
7		6 degree 23	Grad 19 m		степень, 23; градус, 19
8		7 degree of accuracy 7	Genauigkeitsgrad 8 m		степень точности, 7
9		8 extreme 10	extrem 5		крайний, 10
10		9 extreme accuracy 6	Extremgenauigkeit 2 f		крайняя точность, 6
11		10 order 27	Ordnung 28 m		ряд, 27
12		11 order of accuracy 16	Genauigkeitsordnung 9 f		степень точности, 16
13		12 relative 20	relativ 21		относительный, 20
14		13 relative accuracy 4	relative Genauigkeit 4 f		относительная точность, 4
15	+	14 action 164	Handlung 91 f		действие, 164
16		15 elementary 57	elementar 56		элементарный, 57
17		16 elementary variation 14	Elementarhandlung 91 f		элементарное действие, 14
18		17 integral 32	Integral 32 n		интеграл, 32
19		18 integral action 30	Integralhandlung 91 f		интегральное действие, 30
20		19 out 13	Ausgang 12 m		выход, 13
21		20 operation 10	ausser Betrieb 8 m		не работает, 10

Рис. 3. Словарь после обработки В-алгоритмом (основные лексемны отмечены знаком плюс)

ции по использованию и получению начальных данных, применению алгоритмов и оформлению результатов.

Библиографический список

1. Ковалев, И. В. Англо-немецко-русский частотный словарь по системному анализу / И. В. Ковалев, М. В. Карасева; Сиб. гос. аэрокосмич. ун-т. Красноярск, 2004.

2. Ковалев, И. В. Адаптивный алгоритм обучения иностранной лексики на основе лексически связанных компонентов / И. В. Ковалев, В. О. Лесков, М. В. Карасева // Системы упр. и информ. технологии. 2008. № 4 (34). С. 78–82.

3. Ковалев, И. В. Внутряязыковые ассоциативные поля в мультилингвистической адаптивно-обучающей технологии / И. В. Ковалев, В. О. Лесков, М. В. Карасева // Системы упр. и информ. технологии. 2008. № 3.1 (33). С. 157–160.

I. V. Kovalev, V. O. Leskov, E. E. Shukshina

THE PROCESS OF FREQUENCY DICTIONARY BUILDING ON THE BASE OF LEXICAL RELATED COMPONENTS

The procedure of frequency dictionary building on the base of lexical related components and its practical use is considered. The task of frequency multilinguistic dictionary rebuilding for dataware of foreign vocabulary training technique on the basis of lexical related components is defined.

Keywords: multilinguistic adaptive training technology, lexically related component, information-terminological basis, LRC-methodic.

УДК 539.3

Ю. В. Захаров, В. В. Исакова, К. Г. Охоткин

АНАЛОГИЯ ПЕРЕМАГНИЧИВАНИЯ ОБМЕННО-СВЯЗАННОЙ МАГНИТНОЙ СТРУКТУРЫ И ИЗГИБА УПРУГОГО СТЕРЖНЯ СО СЖАТИЕМ¹

Рассмотрено точное аналитическое решение задачи о перемагничивании магнитной структуры с учетом одноосной анизотропии. Получена зависимость средней по толщине слоя намагниченности от постоянного поля. Найдено точное пороговое значение параметра одноосной анизотропии, при котором начинается раскрытие петли гистерезиса. Перемагничивание магнитной структуры аналогично изгибу со сжатием упругого стержня.

Ключевые слова: гистерезис, магнитная структура, нелинейный изгиб.

Задачи о перемагничивании магнитомягкого слоя на магнито жесткой подложке с закрепленным и свободным магнитным моментом на поверхностях в постоянном магнитном поле, параллельном плоскости слоя, решались на основе аналогии с потерей устойчивости эйлера гибкого стержня под действием продольной силы постоянного направления [1]. Уравнение Ландау–Лифшица в статическом случае сводилось к уравнению типа нелинейного маятника и решалось с указанными граничными условиями. В [1] также были найдены распределения магнитного момента и пороги перехода магнитомягкого слоя в веерное состояние в зависимости от магнитного поля, приложенного антипараллельно направлению закрепления.

В работе [2] рассматривалось перемагничивание такого слоя с учетом одноосной анизотропии в плоскости слоя, что привело к уравнению типа нелинейного маятника с дополнительным членом:

$$\alpha \left(\frac{d^2 \varphi}{dz^2} \right) + h \sin \varphi - \beta \sin \varphi \cos \varphi = 0, \quad (1)$$

где φ – угол, зависящий от координаты z по толщине слоя в плоскости этого слоя между вектором намагниченности и осью x , совпадающей с осью легкого намагничивания; α – постоянная обмена ферромагнетика; $h = \text{const}$ – внешнее поле; β – константа одноосной анизотропии, $\beta < 4\pi$.

В [2] было найдено точное решение уравнения (1) в виде дробно-нелинейных комбинаций эллиптических функций, зависящих от одного параметра, определяемого величиной внешнего поля. Эти дробные выражения затем были приближенно представлены в виде полиномов, что позволило найти среднюю по толщине слоя намагниченность и показать наличие гистерезиса при перемагничивании. Порог перемагничивания был найден в виде $h_u + \beta$, где $h_u = (\pi/2)^2 \alpha / d^2$, здесь d – толщина слоя.

Рассмотрим точное решение этой задачи в виде дробно-нелинейных комбинаций эллиптических функций

$$\cos \varphi = \frac{-1 + 2k^2 s^2 - \xi^2 s^2}{1 - \xi^2 s^2}, \quad (2)$$

¹ Работа поддержана программой Минобрнауки Российской Федерации «Развитие научного потенциала высшей школы», (проект 2.1.1/735).